

Confidence Interval Calculation For Specific Subject Test Results

Estimating a confidence interval begins with the assumption that an individual has a true mean breath alcohol concentration denoted by μ . If we were able to obtain an infinite number of measurements and compute their unbiased mean we would obtain this true value. Instead, we are only able to obtain a limited number of measurements, two in the case of breath alcohol analysis, from which we compute their mean - the sample mean denoted by \bar{Y} . This sample mean is an estimate of the true mean μ . Having obtained this sample mean, we are interested in how good this value is as an estimate of the true value μ . The confidence interval quantifies this estimate. The confidence interval provides an interval, symmetric about the sample mean \bar{Y} , from which we infer that we have included the true mean: with some stated level of probability (i.e., confidence). In order to be reliable, the confidence interval estimate must include all sources of uncertainty known to significantly influence the measurement process. The following provides the information and step-by-step computations necessary for determining the 99% confidence intervals for a particular individual's breath alcohol concentration. An actual case from the breath test database is used as an example.

Information Necessary:

1. Datamaster instrument: 949194
2. Date of Test: April 18, 2006
3. Duplicate Test Results: 0.082 and 0.084 g/210L
4. Previous New Simulator Mean Value:
 - a. this is the mean of the first n=5 simulator results following the solution change immediately preceding the subject's test
 - b. in this example, the solution was changed on April 17, 2006 and the first n=5 results were 0.084, 0.084, 0.084, 0.083 and 0.083 g/210L
 - c. in this example, one of the simulator results includes the one associated with the subject's test (not necessary)
 - d. the mean of these results was 0.0836 g/210L
 - e. the solution batch number was 06001
 - f. the new solution is denoted in the database by "NEW/SOL" under the column headed by "Citation"
5. The Toxicology Lab Reference Value for Batch 06001: 0.0826 g/210L
6. Quality Assurance Results
 - a. this occurred on July 20, 2005
 - b. we select the mean results that are nearest to the subject's breath alcohol results, which in this case was 0.0811 g/210L

- c. we note that the systematic error at this concentration was +1.76%
- 7. Quality Assurance Batch and Value:
 - a. the batch was 05019 with a reference value of 0.0797 g/210L
- 8. Plot of the Uncertainty Function along with linear equation
- 9. Tables for the t distribution and the standard normal Z distribution
- 10. All of the relevant documents are available on the Breath Test Section web site at: <http://breathtest.wsp.wa.gov>

The Following Steps Summarize the Calculation of the 99% Confidence Interval

Step 1: Compute sample mean breath alcohol concentration

$$0.082/0.084 \text{ g/210L} \rightarrow \text{mean} = \bar{Y} = 0.0830 \text{ g/210L}$$

Step 2: Adjust the breath alcohol mean by the amount of systematic error determined from simulator results and toxicology lab reference values based on either the Field Results or the Quality Assurance Results.

1. Computing the systematic error based on the first n=5 field simulator results after installing a new solution on April 17, 2006:

$$SE = \left[\frac{\bar{X}_{Sim} - R}{R} \right] \bullet 100 = \left[\frac{0.0836 - 0.0826}{0.0826} \right] \bullet 100 = +1.21\% \quad \text{Eq.1}$$

2. Computing the systematic error based on the results at the time of the Quality Assurance performed on July 20, 2005 using the 0.0797 g/210L reference value we obtain:

$$SE = \left[\frac{\bar{X}_{Sim} - R}{R} \right] \bullet 100 = \left[\frac{0.0811 - 0.0797}{0.0797} \right] \bullet 100 = +1.76\% \quad \text{Eq.2}$$

where: \bar{X}_{Sim} = mean simulator alcohol result
 R = reference value from toxicology lab
 SE = systematic error or bias

Step 3: Summarize the two computed systematic errors are as follows:

Date	Source	Reference	Systematic Error
		Concentration (g/210L)	
07/20/2005	QA	0.0797	+1.8%
04/17/2006	Field	0.0826	+1.2%

Step 4: Identify which estimate of the systematic error should be applied.

Time proximity, concentration and benefit to the subject should all be considered as criteria for this determination. The Quality Assurance results were determined approximately nine months earlier while the field control standard results were determined very near the time of the subject's breath test measurements. Both are reasonably close in concentration to that of the subject results. The field simulator results, therefore, would be a better estimate of the systematic error associated with the subject's results because they are closer in time. However, since the QAP estimate of the bias (+1.8%) favors the subject, we will use this for our corrected estimate. We conclude, therefore, that the instrument is biased high by +1.8% and we adjust the subject mean results accordingly.

Step 5: Adjust the subject's mean results by +1.8% according to the following equation:

$$\bar{Y}_{Corr} = \left[\frac{\bar{Y}}{1 + \frac{\bar{X}_{Sim} - R}{R}} \right] = \left[\frac{\bar{Y}}{1 + SE} \right] = \frac{\bar{Y} R}{\bar{X}_{Sim}} \quad Eq.3$$

where: \bar{Y}_{Corr} = the corrected sample mean

\bar{Y} = the mean of the subject's original results

\bar{X}_{Sim} = the mean of the simulator results

R = the reference value

Incorporating the estimates for the +1.8% bias yields a corrected sample mean of:

$$\bar{Y}_{Corr} = \left[\frac{0.0830}{1 + \frac{0.0811 - 0.0797}{0.0797}} \right] = \left[\frac{0.0830}{1 + 0.018} \right] = 0.0815 \text{ g / 210L}$$

Step 6: Compute the combined uncertainty or standard deviation (S_Y) associated with a single breath alcohol measurement:

The standard deviation is estimated from the uncertainty contributed from four sources: biological/sampling, analytical, water/air partition coefficient for the simulator and traceability. The combined uncertainty (standard deviation) is estimated according to:

$$CV_Y = \sqrt{CV_{Biological}^2 + CV_{Analytical}^2 + CV_{Partition}^2 + CV_{Traceability}^2} \quad Eq.4$$

where: CV = coefficient of variation for each component

$CV_{Partition}^2$ = estimated from literature

$CV_{Traceability}^2$ = estimated from traceability to NIST

The biological and analytical components in equation 4 are combined into one estimate determined from equation 5 which was developed from the analysis of 30,524 field duplicate tests on nearly 200 Datamaster and Datamaster CDM instruments.

$$S_Y = 0.0262 \bar{Y}_{Corr} + 0.00103 \quad Eq.5$$

For the present example, equation 5 yields:

$$S_Y = 0.0262 \bar{Y}_{Corr} + 0.00103 \Rightarrow S_Y = 0.0262(0.0815) + 0.00103 = 0.0032 \text{ g} / 210L$$

Incorporating the estimates into equation 4 yields:

$$CV_Y = \sqrt{\left[\frac{0.0032}{0.0815}\right]^2 + \left[\frac{0.01}{1.22}\right]^2 + \left[\frac{0.0014}{0.0980}\right]^2} \Rightarrow \frac{S_Y}{0.0815} = 0.0426 \Rightarrow S_Y = 0.0035 \text{ g} / 210L$$

Step 7: Compute the 99% confidence interval according to:

$$\bar{Y}_{Corr} \pm t_{1-\alpha/2, df} \frac{S_Y}{\sqrt{2}} \quad Eq.6$$

where: \bar{Y}_{Corr} = the corrected sample mean

t = value from t-table associated with a two-tailed 99% level of confidence and infinite degrees of freedom, 2.576 provides the 99% confidence interval

S_Y = the combined uncertainty or standard deviation estimated in step 6 for a single measurement

Incorporating the estimate from step 6 into the confidence interval of equation 6 we obtain:

$$0.0815 \pm 2.576 \frac{0.0035}{\sqrt{2}} = 0.0815 \pm 0.0064 \text{ g / 210L}$$

The 99% confidence interval in this example would be:

$$0.0751 \text{ to } 0.0879 \text{ g/210L}$$

Estimating the Probability That the True Mean Results Are Above 0.080 g/210L

The following analysis estimates the probability that the subject's true population mean (μ) is greater than 0.080 g/210L. The same approach can be applied to any other concentration of interest. We begin by expressing the confidence interval as a probability statement:

$$P \left[\bar{Y}_{Corr} - Z_{(1-\alpha/2)} \frac{S_Y}{\sqrt{n}} \leq \mu \leq \bar{Y}_{Corr} + Z_{(1-\alpha/2)} \frac{S_Y}{\sqrt{n}} \right] = \pi \quad Eq.7$$

where: π = the probability estimate
 $Z_{(1-\alpha/2)}$ = the variable having the standard normal distribution

Note that the value Z is used here rather than the value t as in equation 6. Since we are assuming an infinite number of degrees of freedom, both values are equivalent. We begin by noticing that we are interested only in the probability that μ exceeds a lower limit and are not concerned about the upper limit. Therefore, we let the upper limit go to plus infinity:

$$P \left[\bar{Y}_{Corr} - Z_{(1-\alpha/2)} \frac{S_Y}{\sqrt{n}} \leq \mu \leq \infty \right] = \pi \quad Eq.8$$

Next, we notice that the lower limit of equation 8 is equivalent to our lower limit of interest, 0.080 g/210L. We set the two equal, introduce our known information and solve for $Z_{(1-\alpha/2)}$:

$$\bar{Y}_{Corr} - Z_{(1-\alpha/2)} \frac{S_Y}{\sqrt{n}} = 0.080 \Rightarrow 0.0815 - Z_{(1-\alpha/2)} \frac{0.0035}{\sqrt{2}} = 0.080 \Rightarrow Z_{(1-\alpha/2)} = 0.61$$

We then rearrange the probability statement of equation 8 and introduce our estimate for $Z_{(1-\alpha/2)}$:

$$P\left[\bar{Y}_{Corr} - Z_{1-\alpha/2} \frac{S_Y}{\sqrt{n}} \leq \mu\right] = P\left[\frac{\bar{Y}_{Corr} - \mu}{SD_Y/\sqrt{n}} \leq Z_{1-\alpha/2}\right] = P[Z \leq Z_{1-\alpha/2}] = P[Z \leq 0.61] = 0.7291$$

From the standard normal tables we see that $P[Z \leq 0.61] = 0.7291$. This corresponds to the statement that: the probability that the individual's true mean breath alcohol concentration is greater than 0.080 g/210L is 0.7291. Conversely, the probability that the individual's true mean breath alcohol concentration is below 0.080 g/210L is 0.2709.

The Assumptions Associated With Estimation of Confidence Intervals:

1. The distribution of within-subject breath alcohol measurements, and thus their sample means, are normal (Gaussian).
2. The computed standard deviation (S_Y) is a valid estimate based on the large number of subject field data, estimates of water/air partition coefficients from the literature and assumed values and estimates from traceability to NIST.
3. The 99% level of confidence is selected for forensic purposes. Other levels can be easily employed by selecting a different value for t in equation 6.
4. The estimate of the biological and analytical components from equations 4 and 5 are probably larger than necessary since they involved thousands of subjects and operators and approximately 200 different instruments.
5. The method of confidence intervals will be robust even for non-normal distributions (e.g., will also include the population mean approximately 99% of the time).
6. Since the population mean (μ) is a fixed but unknown quantity, 99% of the confidence intervals computed from duplicate samples obtained from the subject will include μ .
7. The confidence interval expresses the uncertainty due to sampling variability only, not from any bias in the experimental design or performance.

References

1. Altman, D.G., Machin, D., Bryant, T.N. and Gardner, M.J., Statistics With Confidence: Confidence Intervals and Statistical Guidelines, Second Edition, London: British Medical Journal, 2000.
2. Sokal, R.R. and Rohlf, F.J., Biometry, 2nd ed., New York: W.H. Freeman and Co., 1981, pp. 147-152.
3. Gore, S.M., "Assessing Methods - Confidence Intervals", British Medical Journal, Vol 283, Sept. 5, 1981, pp. 660-662.
4. Fisher, L.D. and van Belle, G., Biostatistics: A Methodology for the Health Sciences, New York: John Wiley and Sons, 1993, p. 305.

Revised 6/20/2007

R. G. Gullberg